

# Towards a Theory of High Confidence Networked Control Systems: Action Webs

Shankar Sastry

Dean and Roy W. Carlson Professor of Engineering  
University of California, Berkeley  
August 9<sup>th</sup>, 2011

Joint work with Saurabh Amin and Galina A. Schwartz  
4th International Symposium on Resilient Control Systems



## Motivation: Cyber-Security

Sensor networks & Networked Control Systems (NCS)  
NCS vulnerabilities

## Cyber-security for NCS

1. Threat assessment
2. Attack diagnosis
3. Resilient control

## Conclusions and ongoing work

## Motivation: Cyber-Security

Sensor networks & Networked Control Systems (NCS)

NCS vulnerabilities

## Cyber-security for NCS

1. Threat assessment
2. Attack diagnosis
3. Resilient control

## Conclusions and ongoing work

# The swarm at the edge of the cloud

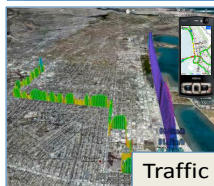
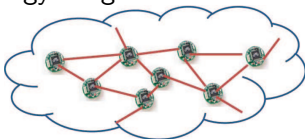




# Ubiquitous instrumentation

## Wireless Sensor Networks (WSN) for infrastructure monitoring

- Environmental systems
- Structural health
- Construction projects
- Energy usage

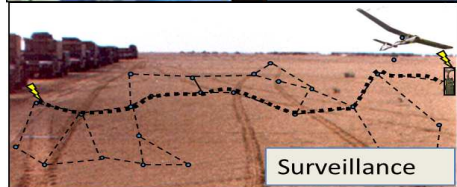


Courtesy: UCB-CEE Systems Faculty

# Sensor webs everywhere

Change detection: Thresholds, phase transitions, anomalies

- Security systems
- Health care
- Wildfire detection
- Fault diagnosis
- Tracking & surveillance



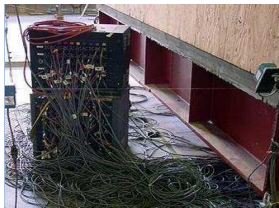
# Widely deployed in critical infrastructures

## Supervisory Control & Data Acquisition (SCADA)

- Robust estimation
  - Noisy measurements
  - Lossy communication
- Real-time control
  - Safety
  - Performance

## COTS IT for SCADA

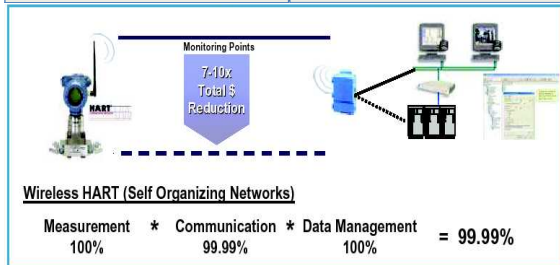
- Cost ↓, Reliability ↑
- Digital and IP based:  
New vulnerabilities!
- Reliability  $\nRightarrow$  Security



Wired networks are costly to maintain



Typical industrial infrastructure ~ \$10B



Source: Emerson case study

# Societal cyber-physical systems

A complex collection of sensors, controllers, compute nodes, and actuators that work together to improve our daily lives

- From very small: Ubiquitous, Pervasive, Disappearing, Perceptive, Ambient
- To very large: Always Connectable, Reliable, Scalable, Adaptive, Flexible

## Emerging Service Models

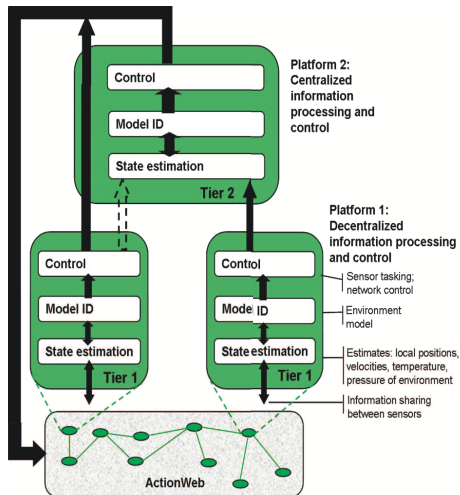
- Building energy management
- Automotive safety and control
- Management of metropolitan traffic flows
- Distributed health monitoring
- Smart Grid

# Action Webs

Observe and infer for planning and modifying action

- Dealing with uncertainty
- Tasking sensors
- Programming the ensemble
- Multiple objectives
- Embedding humans

Example: Building energy management

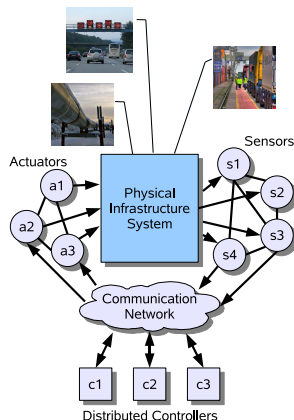


Courtesy: Claire Tomin

# Challenges for Action Webs

## High confidence networked control

- Robust estimation
  - Unreliable communications
  - Mobile sensor & actuator dynamics
  - Distributed parameter systems
- Fault-tolerant networked control
  - Limits on stability, safety, & optimality
  - Scalable model predictive control
- Security & resilience **[Focus of this talk]**
  - Availability, Integrity, & Confidentiality
  - Graceful degradation



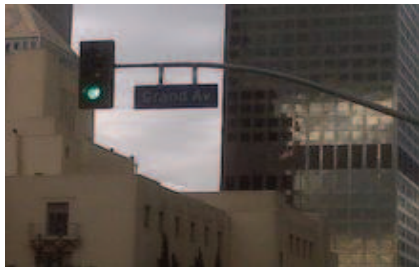
# Cyber-attacks to NCS



Maroochy Shire sewage plant (2000)



Tehama Colusa canal system (2007)



Los Angeles traffic control (2008)



Cal-ISO power system computers (2007)

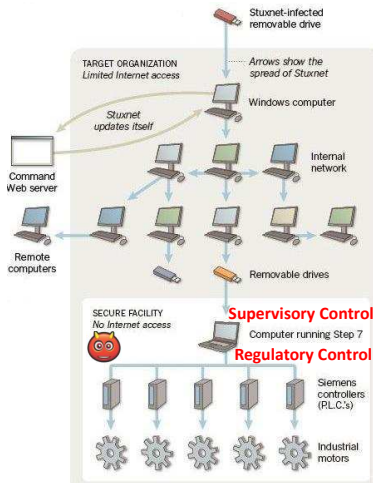
# NCS security concerns

## Attackers

- Malicious insiders
- Computer hackers
  - Cyber criminals
  - Cyber warriors
  - Hacktivists
  - Rogue hackers
  - Corporate spies

## Stuxnet worm

- Targets SCADA systems
- Four zero-day exploits, antivirus evasion techniques, p-2-p updates, network infection routines
- Reprograms *Programmable Logic Controller (PLC)* code



Source: Symantec, NYT



# Previous work in WSN security

- 1 Secure communication
  - SPINS: Security protocols for WSNs ([Perrig, Culler, Tygar](#))
  - TinySec: Link layer encryption ([Karlof, Sastry, Wagner](#))
- 2 Robust aggregation
  - SIA: Secure Information Aggregation ([Przydatek, Song, Perrig](#))
  - Resilient Aggregation ([Wagner](#))
- 3 Sybil Attack
  - Countermeasures ([Newsome, Shi, Song, Perrig](#))
- 4 Secure location verification
  - Verification of location claims ([N. Sastry, Wagner](#))
- 5 Robust localization
  - Statistical methods for robust localization ([Li, Trappe, et.al.](#))
  - SeRLoc ([Lazos, Poovendran](#))
- 6 Cryptographic Key distribution protocols
  - Random key distribution protocol ([Perrig, Song, Gligor](#))

# Previous work in security is not enough

## Missing:

- How is data collected by NCS used?
- Resilient control & anomaly detection for NCS

## System Design

- Least Privilege Principle
- Separation of Duty

## Software Validation

- Correct implementation of system design
- Minimize vulnerabilities and bugs

## Network Security

- End-to-end integrity, confidentiality, availability
- Network intrusion detection

## Device Security

- Trusted Platform Modules (TPM): device integrity

# Cyber-security for NCS

## Classical approaches

- **Cyber**: Computer (IT) security
  - Prevention, detection, and resilience mechanisms
- **Physical**: Robust (fault-tolerant) control
  - Trade-offs: Cost vs. Robustness [to random disturbances]

## Open questions

- Effect of cyber-attacks on control algorithms?
- Faults vs. Attacks?
- Reliability vs. Security?
- Individual vs. Social incentives [to secure]?

Proposal: Robust control + IT security  $\Rightarrow$  NCS security

# Cyber-security for NCS: three problems

## 1 Threat assessment

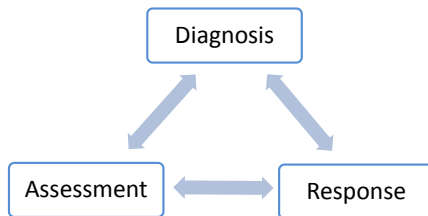
- How to model attacker and his strategy?
- Consequences to the physical infrastructure

## 2 Attack diagnosis

- How to detect manipulations of sensor-control data?
- Stealthy [undetected] attacks

## 3 Resilient control

- Design of resilient control algorithms?
- Incentive mechanisms to improve NCS reliability & security



## Motivation: Cyber-Security

Sensor networks & Networked Control Systems (NCS)

NCS vulnerabilities

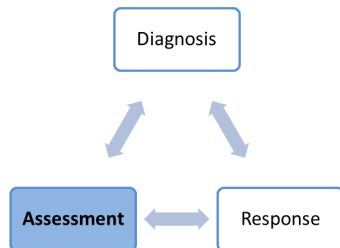
## Cyber-security for NCS

1. Threat assessment
2. Attack diagnosis
3. Resilient control

Conclusions and ongoing work

# Threat assessment

- How to model attacker and his strategy?
- Consequences to the physical infrastructure



Field operational test on the Gignac canal network  
[Amin, Litrico, Sastry, Bayen. HSCC'10]

Models of deception and denial-of-service (DoS) attacks  
[Amin, Cárdenas, Sastry. HSCC'09]

Assessment for Tennessee Eastman process control system (TE-PCS)  
[Cárdenas, Amin, Lin, Huang, Sastry. ASIACCS'11]

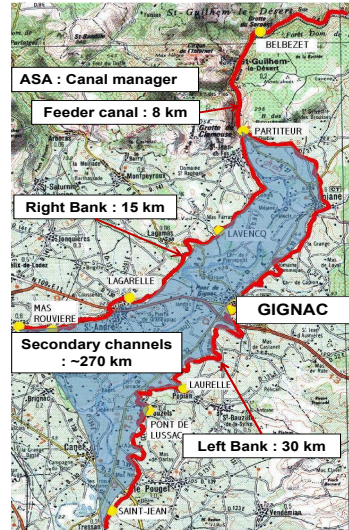
# Gignac water canal network

## SCADA components

- Level & velocity sensors
- PLCs & gate actuators
- Wireless communication
- Multiple stakeholders



Communication station



Map of Gignac canal

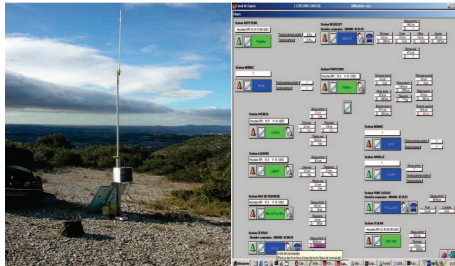
Presented by permission from Cemagref, France

# Gignac canal network

## Physical infrastructure



## Cyber infrastructure





# Reported attacks on water SCADA systems

## Gignac canal system attacks

- Stealing water by compromising sensors
- Tampering PLCs
- Theft of solar panels

## Other SCADA vulnerabilities

- Time between telemetry requests can be used for malicious traffic injection
- Encryption provides confidentiality but does not provide data integrity

### Gignac Le canal victime d'actes de vandalisme à répétition



Depuis le 21 juin, le canal de Gignac est victime d'actes malveillants sur l'ouvrage de l'aqueduc de l'Aurelle (derrière le lagunage de Popian) : effondrement du radier du canal puis dégradation des réparations mises en place (retrait des boulons de serrage, mettant gravement en péril la pérennité de l'aqueduc). L'ouvrage de l'Aurelle permet la continuité du transport de l'eau vers les parcelles du périmètre irrigué situé sur les communes de Pouzols, Le Pouget, Tressan et Puilacher, soit près de 900 ha, pour lesquels l'apport d'eau estival est essentiel. Ces agissements ont fait l'objet de constats par les brigades de gendarmerie et de plaintes contre X. Il est à noter que l'intégralité du patrimoine de l'Association syndicale autorisée du canal de Gignac est un ouvrage public, dont la destruction, la dégradation ou la détérioration peuvent faire l'objet de poursuites et être punies de trois ans d'emprisonnement et de 45 000 € d'amende.

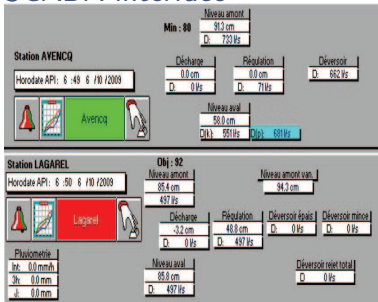
Courtesy: C. Hugodot, Manager

# Regulatory control of canal pools

## Control objective

- Manipulate gate opening
- Control upstream water level
- Reject disturbances (oftake withdrawals)

## SCADA interface



## Avencq cross-regulator



# Defender and attacker models

## Defender

- Estimate Model [Freq. Domain]

$$\hat{y}_i^d = \frac{a_i^d}{s} e^{-\tau_i s} \hat{q}_{i-1} - \frac{a_i^d}{s} [\hat{q}_i + \hat{p}_i]$$

Parameters:  $a_i^d, \tau_i$ , Laplace variable:  $s$

- Design robust decentralized PI control

$$\hat{q}_{i-1} = \kappa_{i-1i} \hat{y}_i^d, \quad \hat{q}_i = \kappa_{ii} \hat{y}_i^d$$

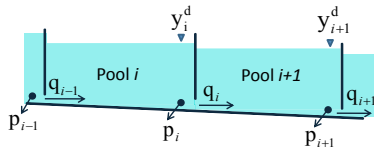
Controllers:  $\kappa_{i-1i}, \kappa_{ii}$

## Attacker

- Compromise  $y_i^d$  and inject  $g_i$

$$\tilde{y}_i^d = y_i^d + g_i$$

- Regulate  $p_i$  to steal water



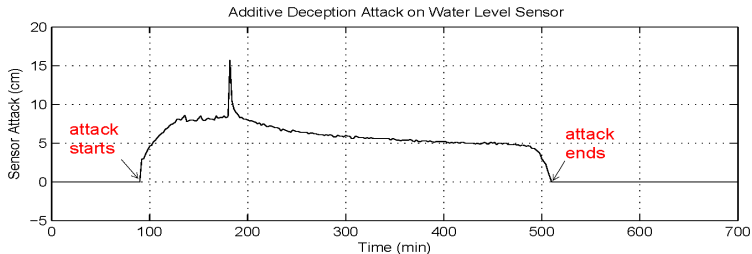
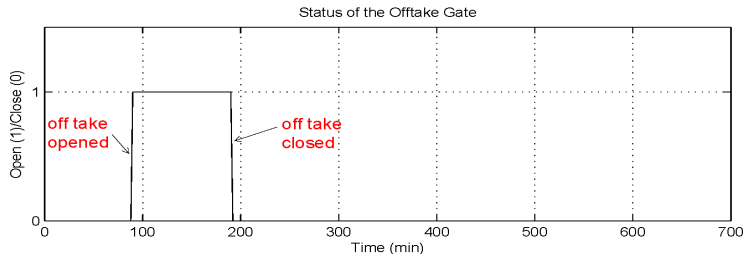
Test site before attack



Test site after attack

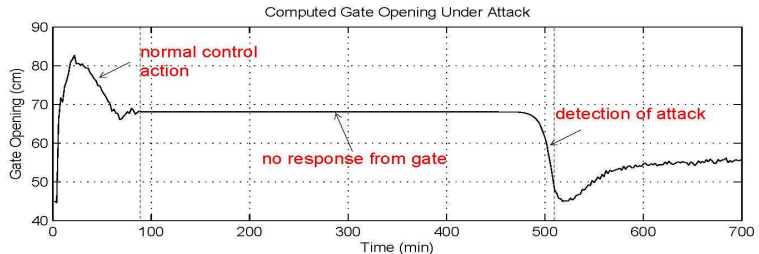
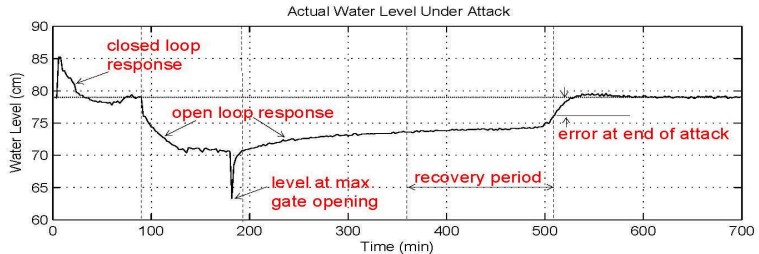
# Cyber-attack on the Avencq canal pool

Field operational test (October 12<sup>th</sup>, 2009)



# Cyber-attack on the Avencq canal pool

## Successful attack



# Cyber-attacks on NCS

## Cyber Attacks

### SCADA Manager [IT Security] **A6**

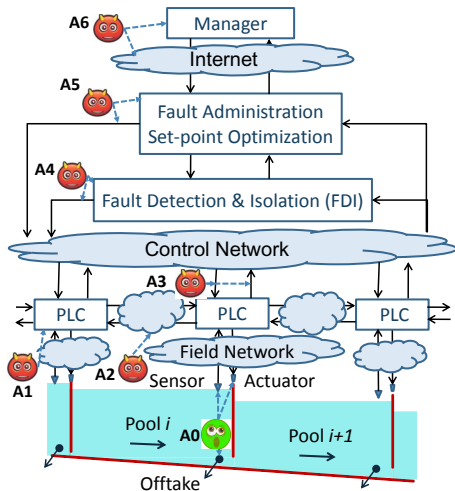
- Unauthorized access, Viruses

### Supervisory Control **A3-A5**

- Deception: set-point change, parameter substitution
- Denial-of-Service (DoS): network flooding, process disruption

### Regulatory Layer **A1-A2**

- Deception: compromise of measurements & controls, spoofing, replay
- DoS: jamming,  $\uparrow$  comm. latency

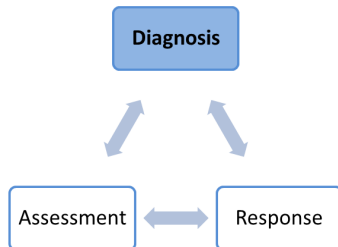


### Physical Faults [Control th.] **A0**

- Sensor-actuator faults
- Unauthorized leaks

# Attack diagnosis

- How to detect manipulations of sensor-control data?
- Stealthy [undetected] attacks



Observer-based diagnosis for Gignac SCADA system  
[Amin, Litrico, Sastry, Bayen. IEEE TCST'11 ]

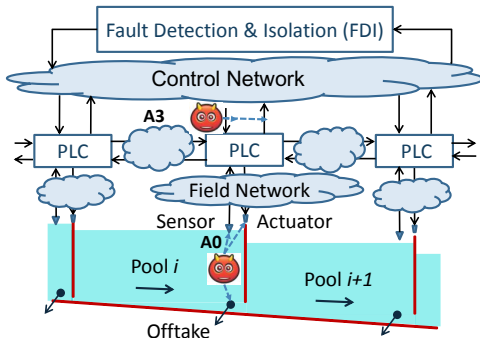
Non-parametric CUSUM statistic based diagnosis for TE-PCS  
[Cárdenas, Amin, Sastry, et.al. ASIACCS'11]

Study of stealthy attacks on power system state estimators  
[Teixeira, Amin, Sandberg, Johansson, Sastry. IEEE CDC'10]

# Attacks on supervisory control layer

## Supervisory Layer Attacks A3

- Deception: set-point change, parameter substitution
- Denial-of-Service (DoS): network flooding, process disruption



## Physical Faults/Attacks A0

- Sensor-actuator faults
- Unauthorized withdrawals

Design of a model-based diagnosis scheme

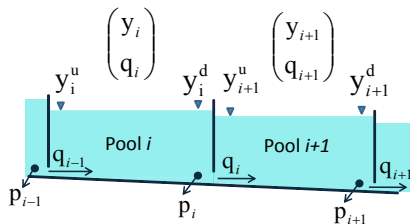


# Flow model

## Linear hyperbolic conservation laws

$$\partial_t \xi_i(t, x) + A(x) \partial_x \xi_i(t, x) + B(x) \xi_i(t, x) = 0,$$

- State:  $\xi_i = (y_i, \quad q_i)^\top$
- Domain:  $x \in (0, l_i), t \geq 0$
- Boundary conditions
  - 1  $q_i(t, 0) = q_{i-1}$
  - 2  $q_i(t, l_i) = q_i + p_i(t)$
- Initial conditions
  - 1  $y_i(0, x) = \bar{y}_i(x)$
  - 2  $q_i(0, x) = \bar{q}_i(x)$



## Variables

### Measurements

- Upstream level:  $y_i^u$
- Downstream level:  $y_i^d$

### Controls

- Upstream discharge:  $q_{i-1}$
- Downstream discharge:  $q_i$

### Disturbances

- Offtake withdrawal:  $p_i$

# Finite-dimensional [approximate] model

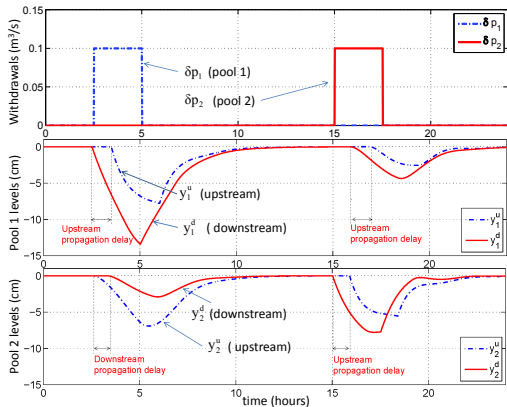
## Delay Differential System

$$\dot{x}(t) = \sum_{i=0}^r A_i x(t - \tau_i) + \sum_{i=0}^r B_i u(t - \tau_i) + \sum_{i=1}^r E_i f_i(t)$$

$$y(t) = Cx(t)$$

For two-pool system:

- State  $x := (y_1^u, y_2^u, y_1^d, y_2^d)^T$
- Input  $u := (u_0, u_1, p_1, p_2)^T$
- Output  $y := (y_1^u, y_2^u, y_1^d, y_2^d)^T$
- Unauthorized withdrawals  
 $f_i(t) := (\delta p_i(t), \delta p_i(t - \tau_i))^T$



# State Estimation

## System

$$\begin{aligned}\dot{x}(t) &= \sum_{i=0}^r A_i x(t - \tau_i) + \sum_{i=1}^r B_i u(t - \tau_i) + E f(t) \\ y(t) &= Cx(t) + H g(t)\end{aligned}$$

- $f$ : unauthorized withdrawals
- $g$ : deception attack on sensors

## Unknown Input Observer (UIO)

$$\begin{aligned}\dot{z}(t) &= \sum_{i=0}^r F_i z(t - \tau_i) + \sum_{i=0}^r T B_i u(t - \tau_i) + \sum_{i=0}^r G_i y(t - \tau_i) \\ \hat{x}(t) &= z(t) + N y(t)\end{aligned}$$

- $F_i, G_i, T, N$ : unknown matrices
- $z$ : observer state
- $\hat{x}$ : state estimate

# Diagnosis scheme for unauthorized withdrawals

## Unknown Input Observer (UIO): design problem

For  $\mathbf{f} \equiv \mathbf{g}$ , find  $F_i$ ,  $G_i$ ,  $T$  and  $N$  such that  $\hat{\mathbf{x}}(t)$  asymptotically converges to  $\mathbf{x}(t)$ , regardless of unauthorized withdrawals  $\mathbf{f}(t)$ .

### Theorem

An asymptotically stable UIO exists if

$$\text{rank} \begin{pmatrix} CE \\ H \end{pmatrix} = \text{rank} \begin{pmatrix} E \\ H \end{pmatrix},$$

& set of delay-dependent linear matrix inequalities are feasible. □

(Amin, Litrico, Sastry, Bayen. IEEE TCST I, II (2011))

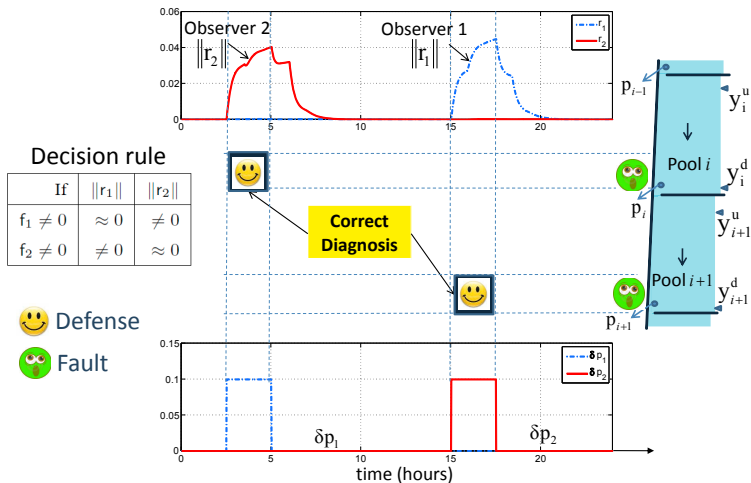
## Diagnosis scheme using the bank of two-observers

Observer residuals  $r_j(t) := y_j(t) - C\hat{x}_j(t)$ ,  $j = 1, 2$

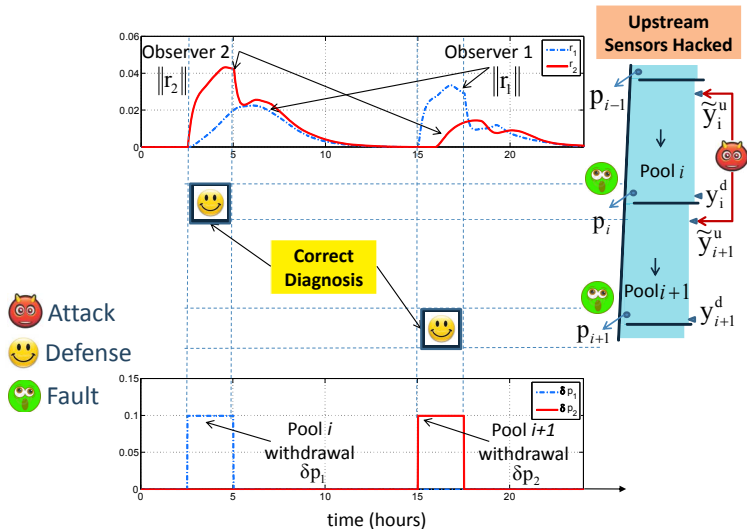
If	$\ r_1\ $	$\ r_2\ $
$f_1 \neq 0$	$\approx 0$	$\neq 0$
$f_2 \neq 0$	$\neq 0$	$\approx 0$

# Diagnosis of unauthorized withdrawals: no attack

Sensors:  $y_i^d, y_{i+1}^d$  and  $y_i^u, y_{i+1}^u$

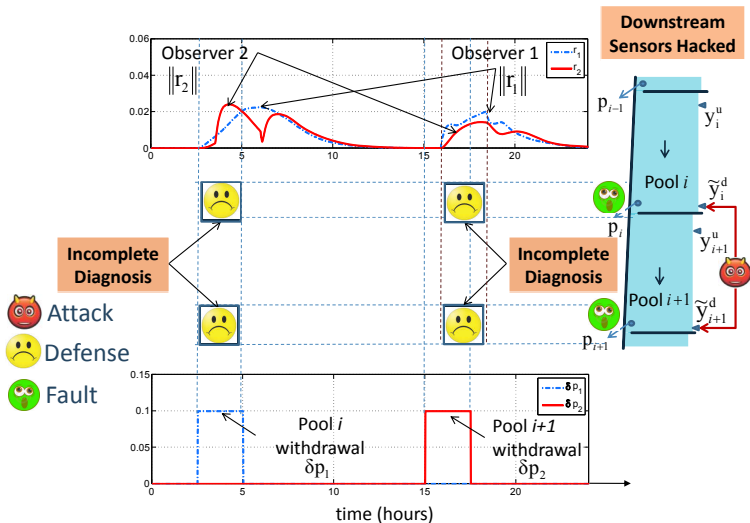


# Attack diagnosis: upstream level sensors hacked



Correct diagnosis of withdrawal in both pools

# Attack diagnosis: downstream level sensors hacked



Withdrawal detected in both pools

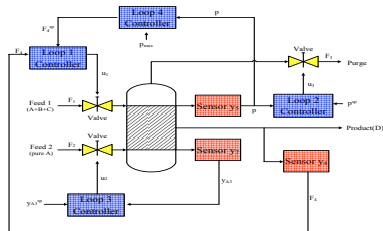
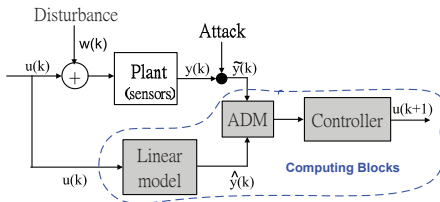
## Recommendations to the European Commission on Canal Automation & the Cemagref Research Institute

- Enhanced model (redundancy) improves detection
- Sensors located closer to the offtakes are critical
- Localized sensor attacks do not lead to global degradation
- Multiple pool sensor attacks can evade detection [stealth]



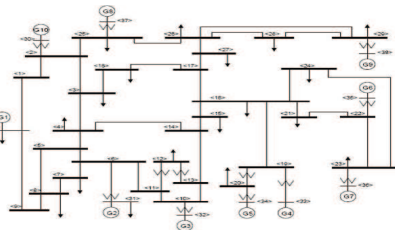
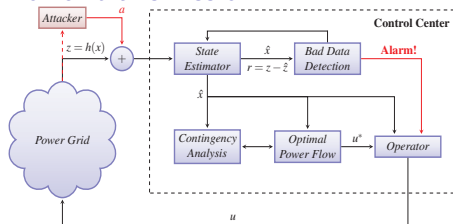
## Attack Diagnosis for [other] SCADA systems

## Process control



[Cárdenas, Amin, Lin, Huang, Sastry. ASIACCS'11]

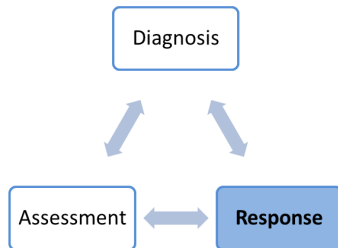
## Power transmission



[Teixeira, Amin, Sandberg, Johansson, Sastry. IEEE CDC'10]

# Resilient control

- Design of resilient control algorithms?
- Fundamental limitations & interdependent security



Stability of hyperbolic PDEs under switching boundary control  
[Amin, Hante, Bayen. IEEE TAC'10]

Incentives to secure under network induced interdependent risks  
[Amin, Schwartz, Sastry. GameSec'10]

Safety-preserving control for stochastic systems under comm. losses  
[Amin, Cárdenas, Sastry. HSCC'09]



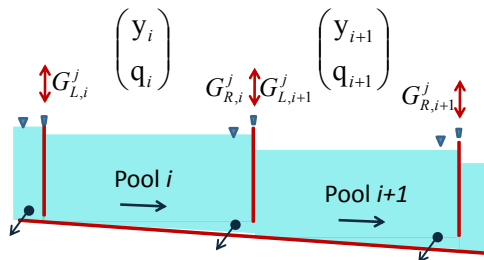
# Attack model: Switching system of PDEs

## Switching attack model

$$\partial_t \xi(t, x) + A^j(x) \partial_x \xi(t, x) + B^j(x) \xi(t, x) = 0, \quad x \in (a, b), \quad t > 0$$

$$\xi_{II}(t, a) = G_L^j \xi_I(t, a), \quad \xi_I(t, b) = G_R^j \xi_{II}(t, b), \quad t \in [0, \infty)$$

$j \in \mathcal{Q}$ , where  $\mathcal{Q} = \{1, \dots, N\}$  is the set of attacker strategies.



Switching attacks: investigation of system stability

# Switching attack: stability

Consider a switching attack  $\sigma(\cdot) : \mathbb{R}_+ \rightarrow \mathcal{Q}$  on the system:

$$\begin{aligned}\partial_t \xi(t, x) + A^{\sigma(t)}(x) \partial_x \xi(t, x) + B^{\sigma(t)}(x) \xi(t, x) &= 0, \quad x \in (a, b), \quad t > 0 \\ \xi_{II}(t, a) &= G_L^{\sigma(t)} \xi_I(t, a), \quad \xi_I(t, b) = G_R^{\sigma(t)} \xi_{II}(t, b), \quad t \in [0, \infty)\end{aligned}$$

## Theorem

Let  $A^j$  be diagonal or pairwise commute, and boundary data satisfy:

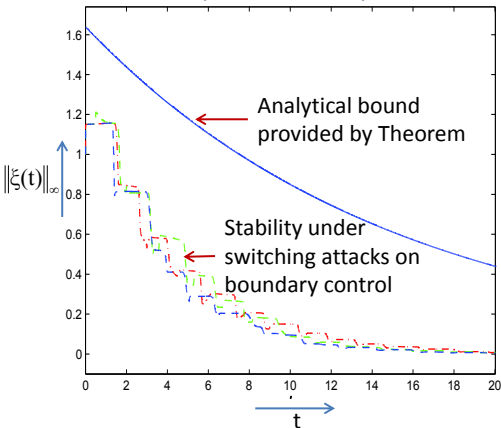
$$\max_{j, j' \in \mathcal{Q}} \rho \left( \begin{bmatrix} 0 & |G_R^{j'}| \\ |G_L^j| & 0 \end{bmatrix} \right) < 1.$$

Then there exists  $\varepsilon > 0$  such that for  $\|B^j(x)\|_\infty \leq \varepsilon$ , the system is exponentially stable under an arbitrary switching attack.  $\square$

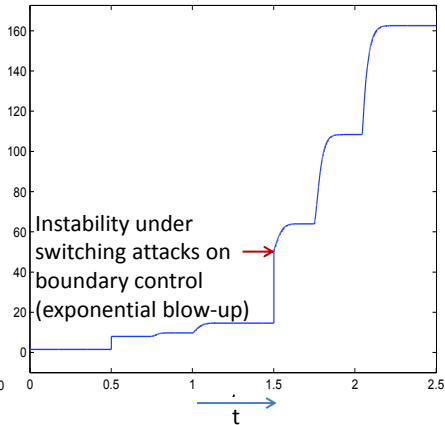
[Amin, Hante, Bayen. HSCC'08, IEEE TAC'10]

# Switching attack: characterization of system stability

All assumptions of stability thm. hold



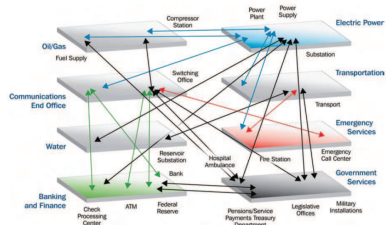
An assumption of stability thm. violated



# Interdependent Security (IDS) & incentives to secure

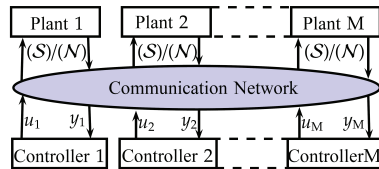
## Security interdependencies due to

- Network induced risks  
⇒ Example: Distributed DOS attacks
- Wide use of COTS IT components  
⇒ Expect increased interdependencies



Infrastructure interdependencies

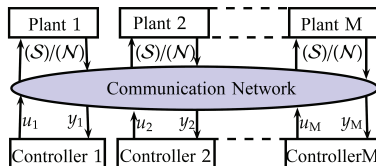
- **Goal:** Security analysis & implementation of control measures
- **Methods:** Game theory & Control theory
- **Observation:** Individual & social incentives differ



Network induced interdependencies

# Interdependent NCS

Two-stage game of plant-controller systems (players)



Each player

- 1 Invests in security [ $V^i = S$  & incurs  $\ell^i > 0$ ] or not [ $V^i = N$ ]
- 2 Chooses inputs  $u_t^i$  for NCS:

$$\begin{aligned}x_{t+1}^i &= Ax_t^i + v_t^i B u_t^i + w_t^i \\ y_t^i &= \gamma_t^i C x_t^i + v_t^i\end{aligned}$$

where  $\gamma_t^i$  &  $v_t^i$  are Bernoulli packet loss processes



# Interdependent failure probabilities

- Failure probabilities:

$$P[\gamma_t^i = 0 \mid V] = \tilde{\gamma}^i(V), \quad P[\gamma_t^i = 1 \mid V] = 1 - \tilde{\gamma}^i(V),$$

- $V := \{V^1, \dots, V^m\}$  Set of player security choices
- Security choices and failure probabilities:

$$\tilde{\gamma}^i(V) = \underbrace{\mathbf{1}_S^i \tilde{\gamma}^i}_{\text{reliability}} + \underbrace{(1 - \mathbf{1}_S^i \tilde{\gamma}^i) \beta(\eta^i)}_{\text{security}},$$

- $\mathbf{1}_S^i$ : Indicator function 1 if  $V^i = S$
- $\eta^i$ : # of insecure players
- $\beta(\eta^i)$ : Interdependence term

$$0 < \beta(\{S, \dots, S, \underbrace{N, \dots, N}_{\eta \text{ players}}\}) < \beta(\{S, \dots, S, \underbrace{N, \dots, N}_{\eta+1 \text{ players}}\}) < 1,$$

# Multiplayer game with interdependent security

- $V := \{V^1, \dots, V^m\}$  Set of player security choices
- $U := \{u_t^1, \dots, u_t^m | t \in \mathbb{N}_0\}$  Set of player control input sequences
- Each player minimizes his total cost:

$$J^i(V, U) = J_I^i(V) + J_{II}^i(V, U),$$

- 1 Security cost

$$J_I^i(V) := (1 - \mathbf{1}_S^i) \ell^i$$

- 2 LQG control cost:

$$J_{II}^i(V, U) := \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} x_t^{i\top} G x_t^i + v_t^i u_t^{i\top} H u_t^i \right]$$

- Social planner minimizes the aggregate cost:

$$J^{\text{SO}}(V, U) = \sum_{i=1}^m J^i(V, U).$$

# Increasing and decreasing incentives to secure

## 2-player game

	S	N
S	$J_{\Pi}^*({S,S}) + \ell^1, J_{\Pi}^*({S,S}) + \ell^2$	$J_{\Pi}^*({S,N}) + \ell^1, J_{\Pi}^*({N,S})$
N	$J_{\Pi}^*({N,S}), J_{\Pi}^*({S,N}) + \ell^2$	$J_{\Pi}^*({N,N}), J_{\Pi}^*({N,N})$

## Increasing incentives

If a player secures, other player gain from securing *increases*:

$$J_{\Pi}^*({N,N}) - J_{\Pi}^*({S,N}) \leq J_{\Pi}^*({N,S}) - J_{\Pi}^*({S,S})$$

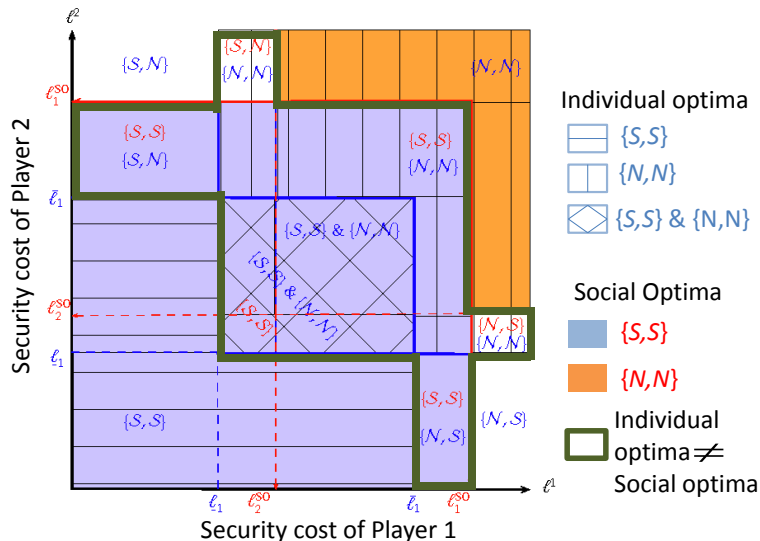
## Decreasing incentives

If a player secures, other player gain from securing *decreases*:

$$J_{\Pi}^*({N,N}) - J_{\Pi}^*({S,N}) > J_{\Pi}^*({N,S}) - J_{\Pi}^*({S,S})$$

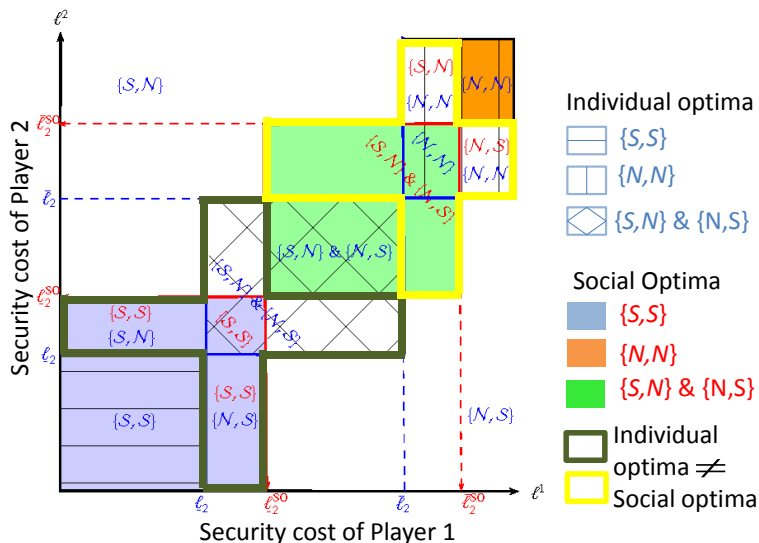
# Individual optima [Nash equilibria] and social optima

## Theorem [Increasing incentive case]



# Individual optima [Nash equilibria] and social optima

## Theorem [Decreasing incentive case]



## Motivation: Cyber-Security

Sensor networks & Networked Control Systems (NCS)

NCS vulnerabilities

## Cyber-security for NCS

1. Threat assessment
2. Attack diagnosis
3. Resilient control

## Conclusions and ongoing work

# Economics of NCS security and reliability

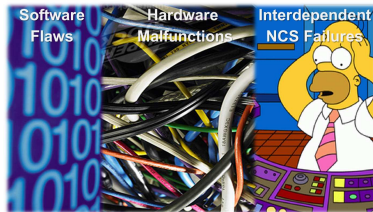
## NCS security & reliability

- Security failures (attacks S) and reliability failures (faults R) are difficult or costly to distinguish
- **Goal:** Model interdependent system failures F

$$\Pr(S \cap R | F) \neq \Pr(S | F) \Pr(R | F)$$

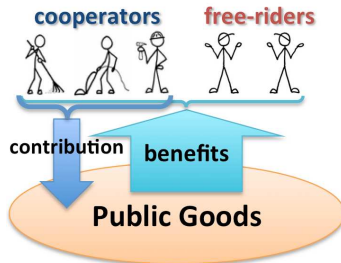
## Negative externalities

- Public goods game
- Information asymmetries
- Property right deficiencies & high enforcement costs
- **Goal:** Develop mechanisms to reduce NCS incentive suboptimality



Courtesy: C. Goldschmidt (Symantec)

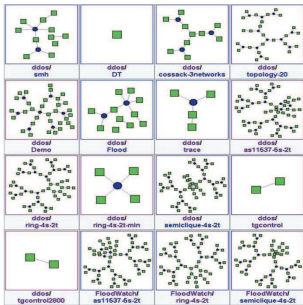
## The Public Goods Game



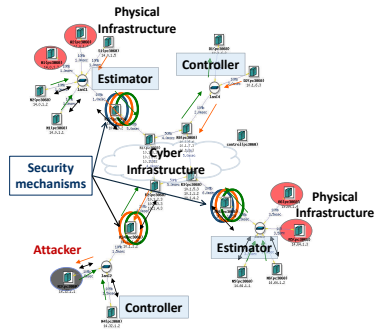
# NCS security experimentation using DETER testbed

## Experiments for networked infrastructure

- Testing
- Validation



Network topologies



Cyber-Security Testbed



cyber-DEfense Technology Experimental Research (DETER) Testbed



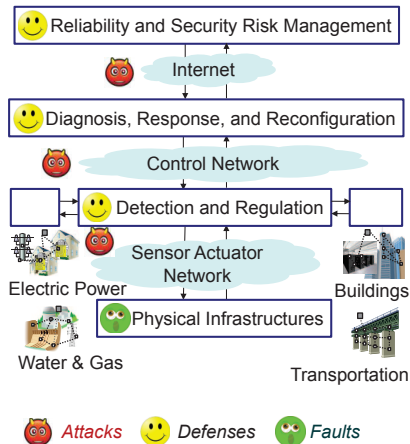
# Towards a theory of high confidence NCS: Action Webs

## Cyber-Security

- Assessment, detection & response
- Stealthy attacks
- Improved diagnostic schemes

## Resilient Control

- Networked and fault-tolerant control
- Fundamental limitations
- Scalable resilient control algorithms
- Incentive mechanisms for security



Thank you for your attention

Shankar Sastry

sastry@coe.berkeley.edu

Visit <http://www.truststc.org> for more information